DOCUMENT RESUME

ED 334 242 TM 016 778

AUTHOR Busch, John Christian; Taylor, Raymond G.
TITLE The Use of Linear (Goal) Programming in the Construction of a Test Blueprint.
SPONS AGENCY North Carolina Univ., Chapel Hill. General Administration.; North Carolina Univ., Greensboro.
PUB DATE 90
NOTE 15p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Cognitive Processes; College Mathematics; Computer Assisted Testing; Education Courses; Higher Education; *Linear Programing; Mathematics Tests; Statistics; *Test Construction; Test Content
IDENTIFIERS *Goal Programming; *Test Specifications

ABSTRACT

The use of a variation of linear programing (goal programing) to develop a test blueprint with multiple specification requirements, is the topic of this paper. The computer program STORM was used to develop a table of test specifications for a test that would measure achievement in introductory statistics in the subdomains of frequency distributions and graphs, central tendency, and variability. Specific goals were minimal length, reasonable testing time, and feasible development cost (in terms of time). Twenty-five specific constraints were specified with regard to cognitive processes and content measured. The developed blueprint could not achieve all the goals, but did allow a compromise reflecting the individually prioritized goals. Goal programing appeared to be a practical method for specifying the dimensions of a test blueprint. Five tables illustrate the goal programming process. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED334242

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JOHN CHRISTIAN BUSCH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Use of Linear (Goal) Programming in the Construction of a Test Blueprint

John Christian Busch
The University of North Carolina at Greensboro

and

Raymond G. Taylor
North Carolina State University

This research was supported in part by a grant from the General Administration, The University of North Carolina and the Collegium the University of North Carolina at Greensboro

TM016778

BEST COPY AVAILABLE

The development of a table of specifications (test blueprint) has important multiple functions. It serves as a guide to the development of a pool of items which reflect the test author's understanding of the construct to be measured. Once a test form has been developed, the blueprint may be used to evaluate whether the particular combination of items selected for the form match the operational description of the trait domain as it is expressed in the blueprint.

The test developer has traditionally considered two dimensions in constructing the blueprint. They are " ... a dimension of content expressing what has been taught and a dimension of process indicating what the examinee is expected to do with that content" (Thorndike, 1982). The content dimension is a function of the particular trait to be measured. For example, in an achievement test of introductory measurement knowledge and skills the content dimension might include general topics such as: test scores, reliability, test validity and so on. The second dimension reflects the cognitive process used by the respondent to respond to the demand of the particular item. A commonly used taxonomy used to describe the second dimension is Bloom's (1956) Taxonomy of Educational Objectives.

The table of specifications is constructed as a two-way table with the content and process dimensions separately arrayed on the vertical and horizontal axes. Crocker and Algina (1986) point out that "...specifications should also indicate the balance that should be maintained in trying to tap these different processes for the test as a whole" (p. 73). A number which is assigned to each cell of the blueprint reflects the importance of the particular process-content combination and is reflected in the number of items and the assigned weight. "In an achievement test these weights usually reflect the amount of time devoted to mastery of material at that level during the course of instruction" (Crocker and Algina, 1986). The need to insure adequate precision of part and total scores will also determine the minimum number of items on the test.

Other internal and external factors must be considered in specifying the number of items. Millman and Greene (1990) state that "In most test settings, the most important of these external factors are the practical or psychological constraint of an available testing time" (p. 338). Thorndike (1982, p. 15) observes that "...the pressure for brevity in testing is very

real. But the pressure for precision is equally real, and the test maker is in a perennial conflict between producing a test that is brief enough to be acceptable to users and creating one that provides enough information to permit useful decisions". Millman and Greene (1989, p. 338) indicate that a major task of the test planner is achieving some sort of balance between these and other "often conflicting demands on test content and form. The proficient test developer must decide what is discretionary and what cannot be compromised, determinations made largely on the basis of intended test purpose"

How best to achieve a solution to a problem when there are several potentially competing objectives or goals? If it is not feasible to achieve all goals, is it possible to accomplish a sub-set of those goals which conform to the test developer's assessment of their relative importance? Can a solution be found in which the extent of failure to meet certain goals is minimized? The management sciences have developed linear programming as one approach to solving such complex problems. "... (M)anagers or decision makers often face problem situations in which more than one objective exists. Goal programming was developed in the 1960's as a procedure for handling multiple objective situations within the general framework of linear programming. Each objective is viewed as a 'goal'. Then given the resource limitations or constraints, the manager attempts to develop decisions that provide the 'best' solution in terms of coming as close as possible to reaching all goals" (Anderson, D., Sweeney, D. and Williams, T., 1979, p. 225). The differences between goal and linear programming are that there are multiple goals (in the goal programming context), that the optimal solution always involves minimization of an objective function and that deviation from goals are minimized in a priority order in goal programming (Taylor, 1990, p. 41). Applications of such approaches in educational research and measurement are not extensive, however recent important work by Adema and van der Linden (1989) and Adema (1990) describes the use of linear programming approaches to test construction. They utilized linear programming models to select items from a pool in order to maximize reliability under various practical constraints, including test administration time and distribution of content. Their work utilizes some important new applications from the management sciences to the solution of measurement problems.

Purpose

This paper reports on an application of goal programming to the solution of a common measurement problem, the construction of a table of specifications for an achievement test to be used in an introductory applied educational statistics course. In preparing test specifications the test designer must deal with several competing goals. First, the test should utilize a minimum number of items in order to achieve a reasonable degree of precision. The items must be distributed across the dimensions of the test in sufficient numbers and in appropriate proportions. At the same time there is a need for the test to be completed within a reasonable period of time. Finally, the cost (in time or dollars) of producing the test should be realized within certain limits. These goals and constraints are potentially in competition with each other; for example, a test which is "long enough" for reliability requirements yet not "too long" is desired.

Method

The problem will be solved with linear (goal) programming using the computer program STORM , version 2.0.

The Achievement Test: The test specifications which will be constructed describe a test in applied educational statistics which measures achievement in three subdomains (frequency distributions and graphs, central tendency, and variability). The items on the test require achievement to be demonstrated by recognition, by comprehension or through the ability to analyze. The overall form of the table of specifications for the hypothetical achievement test is presented in Table 1.

Insert Table 1 about here

The specification of the parameters and their values for a solution in any particular case would be based on theoretical as well as empirical determinations. For example, subjects' time to complete specific types of items might be available from prior experience in assessing the domain, or other sources might be helpful. For example, Tinkelman (1971, p.72) describes the time to complete various kinds of items (vocabulary,

multiple choice, etc.). The theoretical understanding of the construct and the course of instruction would provide a basis for determining the relative importance of different items.

We wished to construct a test blueprint which achieved the three specific goals which are listed in Table 2. Minimal test length was considered to be the most important goal, followed (in order) by reasonable testing time and feasible cost for test development. An ideal solution would achieve all three goals; failing that, the technique should attempt to satisfy goals 1 and 2. If that were not possible, then goal 1 should be satisfied.

Insert Table 2 about here

In addition, the test and blueprint had to satisfy the following constraints: The test should contain both minimum numbers and relative proportions of items which measure different cognitive processes (recognition, etc.) and which assess different content dimensions (frequency distributions and graphs, etc.). Every process-content combination (cell) in the blueprint should contain at least one item. These constraints are stated in detail in Table 2.

Further, it was assumed that each type of item (either classified by content or process) demanded a different cost to the person taking the test (in minutes spent answering the item) and in item development time (minutes spent writing the item based on an item shell). These costs are specified in Table 3.

Insert Table 3 about here

Results and Conclusions

Table 4 contains the objective function which was minimized for this problem and lists the equations which defined the constraints which were satisfied in its solution. The 25 constraints listed in Table 4 are the same constraints described verbally in Table 2. The same numbers are used to identify each.

Insert Table 4 about here

Several of the constraints require explanation. First, the values Xij which are solved represent the number of items in each of the cells of the table of specifications. Each value of Xij is defined in Table 1. D1U , D2V, and D3V are deviational variables which are minimized in the objective function (Table 4). The numeric value in each variable name signifies the relevant goal (ie. goal 1, 2, or 3). The "U" signifies that the goal is under-achieved in the solution and the "V" indicates that the goal is over-achieved. For example, D1V represents over-achievement of goal 1 (more than 40 items). Note that it is not minimized in the function but is allowed to vary (see the constraint statement of goal 1; more than 40 items are possible if that is necessary to achieve a solution. That is certainly consistent with our intent since we wished to have at least 40 items in the test form. In contrast, D1U represents under-achievement of the same goal (less than 40 items on the test). It is included in the objective function; however in the initial solution this variable is set to 0 (constraint #10) so that there is no under-achievement of goal 1 (there may not be less than 40 items on the test form).

As another example, D3V represents over-achievement of goal 3 (total time to develop the test items will greater than 1200 minutes). D3U represents under-achievement of goal 3 (test time less than 1200 minutes). Since D3V is initially set to 0 (see constraint #12), the number of total minutes to develop items is not allowed to exceed 1200 minutes. For the initial solution (see constraint #3) the total number of minutes to develop items for the test is allowed to be equal to or less than 1200 minutes but it may be no greater than 1200 minutes because variable D3V has been set to 0.

A solution to all three goals is sought by attempting to solve the objective function with all of the constraints included in Table 4. If a solution is "not feasible" it means that the three goals may not be simultaneously satisfied. The programmer may then sequentially release the constraints associated with the goals having lower priority. Therefore, it is possible to achieve a solution for all three goals, for the two of the three that are more important, or the one goal of the three which is most important. Of course, it is also possible that no satisfactory solution can be found which meets even the most important goal.

In this problem, a solution satisfying all three goals was not feasible. Goal 3 was regarded as least important and therefore constraint #12 was removed and the program re-run. This constraint (D3V=0) did not allow goal three to be over-achieved and its removal allowed more than 1200 minutes to develop the exam. That solution was feasible. However if it had not been feasible, then the zero constraint associated with goal 2 would have been relaxed, and another solution sought.

The results of the optimal solution for this problem, with goal 3 relaxed, are summarized in Table 5. The final test form in this solution will consist of 60 items allotted to the cells of the three-by-three matrix as indicated. The solution satisfies goal 1; that is, there are at least 40 items. The 60 items on the test will require a total of 156 minutes to complete, which satisfies goal 2. Student testing time is 24 minutes less than the maximum 180 minute limit. This estimate is based on the expected student "cost" (response time) for different types of items as described in Table 3. Goal 3 was not achieved. The 60 items will require 1275 minutes to develop, based on the "cost" of different types of items specified in Table 2. While it was not possible in this situation to achieve all of the goals the solution allows us to achieve a compromise which reflects our individually prioritized goals.

Insert Table 5 about here

Discussion

Our paper explored the use of a variation of linear programming (goal programing) to develop a test blueprint when multiple requirements were placed on its specification. Such solutions are often accomplished rather inefficiently through trial and error. In this application, we found goal programming to be a practical method which may be used to specify the dimensions of a test blueprint.

In this example, the blueprint was designed using goal programming methods and it was assumed that item development would follow at some later point. Under such circumstances, it must be assumed that the estimates utilized (such as the time for students to complete certain item types) are reasonable. We believe that it will be possible to make

reasonably accurate estimates when developing a test in a familiar domain; pilot testing might also provide a basis for such estimates. While we utilized a small number of constraints to illustrate these methods, the approach can be easily modified to accomodate a variety of additional constraints of greater complexity than those of the present example.

We believe that the management sciences offer powerful solutions to many of the problems faced by measurement professionals. For example, Adema (1989;1990) has made significant contributions in utilizing linear programming to solve measurement problems. However, at the present time the measurement literature describing such applications seems to be rather limited. We encourage experimentation to identify other problems which might be solved using these approaches.

References

Adema, J. & van der Linden, W. (1989). Algorithms for computerized test construction using classical item parameters. Journal of Educational Statistics, 14(3), 279-290.

Adema, J. (1990). The construction of customized two-stage tests. Journal of Educational Measurement, 27(3),241-243.

Anderson, D., Sweeney, D. & Williams, T. (1979). An introduction to management science. St. Paul: West.

Bloom, B. (1956). Taxonomy of educational objectives, handbook 1: The cognitive domain, NY: McKay.

Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt.

Emmons, H., Khot, C., Mathur, K. (1989). STORM: Quantitative modeling for decision support. Oakland: Holden Day.

Taylor, R. (1990). Management operations in education. Unpublished document, Greenville, NC: East Carolina University.

Thorndike, R. L. (1982). Applied psychometrics. Boston: Houghton Mifflin.

Millman, J, & Greene, J. (1989) The specification and development of tests of achievement and ability, pp. 335-366. In R. Linn (ed.) Educational measurement (3rd edit) NY:Macmillan.

Tinkelman, S. T. (1971). Planning the objective test. In R.Thorndike (ed.) Educational measurement (2rd edit), Washington, DC:American Council on Education..

Table 1.
Table of specifications for an achievement test in applied educational statistics.[1]

Content	Process:		
	Recognition	Comprehension	Analysis
Frequency	X_{11}	X_{12}	X_{13}
Central Tendency	X_{21}	X_{22}	X_{23}
Variability	X31	X32	X_{33}

[1] X_{ij} is the number of items in a cell

Table 2
Goals and necessary constraints to determine of the number of items in a table of specifications for an achievement test in educational statistics.

Goals:
1. The test should contain a minimum of 40 items.
2. The test should be completed within one three-hour class period (less than 180 minutes).
3. Test development should be of reasonable cost; at $1.00 per minute of development time it should be possible to write the items in less than 20 hours and cost less than $1200.

Constraints:
4. The minimum number of recognition items is 12.
5. The minimum number of comprehension items is 10.
6. The minimum number of analysis items is 15.
7. The minimum number of items assessing frequency distribution is 10.
8. The minimum number of items assessing central tendency is 15.
9. The minimum number of items assessing variance topics is 15.

Three variables (10-12) are used to control the order of solution of the three goals and are defined in Table 4.

The relative number of items at different process levels:
13. There should be at least a 2:1 ratio of comprehension items to recognition items.
14. There should be at least a 2:1 ratio of analysis items to recognition items.

The relative number of items assessing different content:
15. The number of items assessing knowledge of frequency distributions should be less than or equal to the number of items assessing knowledge of central tendency.
16. The number of items assessing knowledge of frequency distributions should be less than or equal to the number of items assessing knowledge of variability.

17. - 25. Each cell in the table will contain at least one item. all Xij >= 1

Table 3
Assumed costs to develop individual items (by content area) and costs in student response time (by process level) for the achievement test.

Student response time per item:		Development cost per item:	
Recognition items=	1 min.	Frequency items =	15 min.
Comprehension items=	3 min.	Cen. Tendency items=	20 min.
Analysis items=	3 min.	Variability items=	30 min.

Table 4
The objective function, goals and constraints used in the linear programing solution.

The objective function minimizes Z

Z= 3D1U + 2D2V + 1D3V
(goal 1) (goal 2) (goal 3)

Goals

1. Goal 1: X11+ X12+ ... +X33+ D1U- D1V = 40 items
2. Goal 2: 1X11+ 1X21+ 1X31+ 3X12+ 3X22+ 3X32+ 3X13+ ...+ 3X33 + D2U- D2V = 180 minutes
3. Goal 3 : 15X11+ 15 X12+ 15X13+ 20X21+ ...+ 30 X31+ ... + 30 X33 + D3U- D3V= 1200 minutes

Other Constraints

Minimum number of items per topic and process level:

4. X11+ X21+ X31 >= 12
5. X12+ X22+ X32 >= 10
6. X13+ X23+ X33 >= 15
7. X11+ X12+ X13 >= 10
8. X21+ X22+ X23 >= 15
9. X31+ X32+ X33 >= 15

Controls order of solution of goals (to be sacrificed in in reverse order if required to obtain a feasible solution):

10. D1U = 0 11. D2V = 0 12. D3V = 0

Relative number of items at different process levels (2 or more:1):

13. -2X11 -2 X21 -2 X31 +X12 +X22 +X32 >= 0
14. -2X11 -2 X21 -2 X31 +X13 +X23 +X33 >= 0

Relative number of items at different content levels (1 or more:1):

15. -X11 -X12 -X13 +X21 +X22 +X23 >=0
16. -X11 -X12 -X13 +X31 +X32 +X33 >=0

Minimum number of items per cell:

17. - 25. all Xij >= 1

Table 5
Solution to the problem: a test blueprint which satisfies goals one and two within the specified constraints

Content	Process: Recognition	Comprehension	Analysis	Total
Frequency	X_{11}= 4	X_{12}= 1	X_{13}= 10	15
Central Tendency	X_{21}= 7	X_{22}= 22	X_{23}= 1	30
Variability	X_{31}= 1	X_{32}= 1	X_{33}= 13	15
Total	12	24	24	60

Note:

1. Goal 1: X11+ X12+ ... +X33+ D1U- D1V = 40 items
(Solution: 60 + 0 - 20 = 40)

2. Goal 2: 1X11+ 1X21+ 1X31+ 3X12+ 3X22+ 3X32+ 3X13+ ...+ 3X33
+ D2U - D2V = 180 minutes
(Solution: 156 +24 - 0 = 180)

3. Goal 3 : 15X11+ 15 X12+ 15X13+ 20X21+ ...+ 30 X31+ ... + 30 X33
+ D3U- D3V= 1200 minutes
(Solution: 1275 + 0 - 75 = 1200)